

Počítačové zpracování slovo tvorby



Počítačové nástroje pro češtinu
Jaro 2020

Adriana Válková

Slovotvorná analýza

- tj. segmentace slova na: 1. (slovotvorný) základ, 2. (slovotvorný) formant
- **výstupem**: A. získání základového slova (*les*), B. slovotvorného postupu a způsobu, které vedly ke vzniku „nového“ slova (postup: prefixálně-sufixální (*pra-*, *-ík*); způsob: derivace)
- **slovotvorná a morfologická analýza**: 1. slovotvornou analýzu má smysl dělat pouze u odvozených slov, 2. morfologická analýza se provádí po slovotvorné analýze (morfologický rozbor základu a formantu)



Struktura (segmentace) slova

- **slovotvorný základ:**
 - celé základové slovo: *do-dělat* > *dělat*
 - slovní tvar: *zmizel-ý* > *zmizel*
 - kmen: *sladk-ost* > *sladký*
 - kořen: *slad-it* > *sladký*
- **slovotvorný formant** (vše, čím se odvozené slovo odlišuje od základového)
 - **afixy:** prefixy (*do-dělat*), sufixy (*děl-ník*), infix (*mal-il-inký*), interfix (*hlad-o-mor*), postfix (*kdo-si*)
 - spoluformanty: změna tvaroslovné charakteristiky (*šikm-o* ↔ *šikm-ý*)
 - průvodní formanty: hláskové alternace – při sufixaci (*pražský* > *Praha*)

Slovotvorné způsoby a postupy

- **základní slovotvorné způsoby**: derivace (odvozování v užším smyslu), kompozice (skládání), abreviace (zkracování)
- **základní slovotvorné postupy**: afixy jsou nositelem slovotvorných postupů (prefixace, sufixace, ...)

Většinou je základové slovo formálně jednodušší (tj. kratší) než slovo (z něj) odvozené, ale ne vždy: desufixace (*honit* > *hon*), deprefixace (*útes* > *tes*)

Dokulil: onomaziologická teorie slovo tvorby

- stále převládající slovo tvorná teorie v českém prostředí
- při odvozování (v širším smyslu) dochází k významovým změnám:
 - transpozice (změna slovního druhu bez významové změny):
např. *starý* > *stáří*
 - modifikace (obměna významu): *mrak* > *mráček*
 - mutace (změna slovního druhu + významu): *vrtat* > *vrtačka*

Počítačové zpracování jazyka – slovotvorba

uvažujme slovo *učitel*

- co vidíme my: slovo s významem...
„učit“



související se slovesem

- co „vidí“ počítač (program): *u-č-i-t-e-l*: 1. znak „u“, 2. znak „č“, ...
- algoritmický popis jazyka (slovotvorby) → co má počítač vidět;
algoritmus = postup (popis kroků)



Nástroje pro počítačové zpracování slovo tvorby (derivace)

- Deriv (Derivancze): seznamy n-tic (Brno)
- DeriNet: vizualizace derivačních vztahů (Praha)
- Morfio: seznamy n-tic (Praha)
- (Ajka): pouze základní derivační vztahy (Brno)

Deriv, Morfio

- derivace (produktivita slovotvorného typu = míra, kterou se podílí na současném tvoření nových slov)
- vyhledání formálně blízkých dvojic (n-tic): základové slovo – odvozené slovo
- vyhledávání pomocí vzorů
- Deriv (deb.fi.muni.cz/deriv), Morfio (<https://morfio.korpus.cz>)
- Deriv: data z korpusu SYN 2000 + czTenTen12, Morfio: data z korpusů (SYN2005, SYN2010, SYN2015)

Úkol: vyhledat...

- z neuter zakončených -o (typu *město*) se tvoří více adjektiva -ový (*dominový*) nebo -ěný (*dřevěný*)?
- z čeho je odvozeno substantivum *knihovna*?
- vzniklo sloveso *černat* z adjektiva *černý* nebo naopak?
- jaké existují prefigované tvary od slovesa *blbnout*?
- kolik derivačních kroků předchází adverbiu *blbečkovsky*?